DESCRIPTION

HIGHLY EFFICIENT METHOD OF GENOME SCANNING

Technical Field

This invention relates to a method for performing high-efficient electrophoresis of multiple nucleic acid samples to detect nucleic acids of interest. The invention also relates to an electrophoresis apparatus used in the method. The method of this invention is useful particularly in detecting polymorphism in genomic DNAs, genetic analysis, genetic mapping, and constructing a contig or a physical map that covers the entire genome of an organism.

Background Art

To detect differences of organisms, such as those between breeds, at the nucleic acid level, techniques such as RFLP (Restriction Fragment Length Polymorphism) and RAPD (Randomly Amplified Polymorphic DNAs) have been conventionally used. In the RELP method, however, large quantities of DNA samples are required, and also genomic maps based on existing RELP markers for the tested organism are necessary to detect a marker proximal to a particular gene. In addition, construction of the map requires substantial time, cost and manpower. Further, only limited organisms have genetic maps that contain RELP markers at sufficient densities. The detection of polymorphism with the RAPD method, in contrast, may be applied to a relatively large number of samples. However, the number of bands stably obtained at a time is limited. Loosening PCR conditions in attempt to increase the number of bands tends to deteriorate the reproducibility of the resulting polymorphic bands.

The AFLP (Amplified Fragment Length Polymorphism) method has been increasingly used because of its ability to compare a large number (50 to 100 or more) of bands at a time, low consumption of DNA samples, and high reproducibility of resulting bands. However, in its original procedure, the sequencing gel is as large as 40 to 50 cm and the nucleic acid bands are detected by autoradiograph using isotopes. Thus, the method requires extensive experience, may be used only in limited

conditions, takes time for detection, and is not capable of analyzing a large number of samples (capable to process only up to several dozen samples at a time).

Recently a method that uses PCR with fluorescent primers and automatic sequencers has been developed to make the band detection easier. However, these types of sequencer are very expensive (¥10,000,000 to ¥20,000,000 or more), and an experiment of this method would occupy the sequencer, which is basically for gene sequencing, for a considerable period of time. Moreover, since the band detection in this method assumes that the procedure is performed with a system using an automatic sequencer, the band of interest cannot be isolated for analysis after the detection step. Thus, there is a major problem that the detected bands cannot be readily detected as SCAR (Sequence Characterized Amplified Region) markers by specific primers. In addition, only 8 to 9 types per set of fluorescent primers are currently available, allowing merely 64 to 81 combinations of primer pairs at most.

A known method for detecting polymorphic bands over an entire genome at a time is RLGS (Restriction Landmark Genome Scanning). However, this method also uses radioisotope and requires several days to detect markers of small quantities. Further, it involves handling of a very large gel (40 x 30 cm) or a long, narrow agarose gel for every sample, requiring extensive experience as well as muscular strength. In addition, cleavage of nucleic acids by restriction endonuclease in the agarose gel requires a large amount of expensive restriction endonucleases, making the procedure costly.

In RLGS on rice genome (450 MB), for example, only limited 8-base restriction endonucleases, such as NotI, may be used to define the labeled portion. In addition, the theoretical upper limit for the number of dots obtained from one electrophoresis cycle is 450 MB $\div$ $4^8 = 13,700$, and in practice, because of the nature of the electrophoresis, the number is 1/3 to 1/6, i.e. 2,000 to 4,000 dots. Moreover, since sample from each individual is electrophoresed on a separate gel, electrophoresis patterns of different samples often do not match completely. Thus, an expensive, large-scale scanner and two-dimensional electrophoresis software are necessary to compare

different individuals.

In the early stage in the construction of genomic libraries in which contiguous clones covering an entire genome were organized and linked based on overlaps between the clones, there were attempts to utilize markers in existing maps such as the RFLP map. However, even in those called high-density maps there are only about 2,000 markers. Even in rice, which has a small genome (450 MB), the average density is merely 200 KB/band or more. Such densities are too low to construct a contig covering the entire genome. On the other hand, the search for such a number of RFLP markers would require vast amount of cost, manpower, and time.

A method used often recently to construct a contig covering an entire genome is as follows: Component clones of a library are cleaved by appropriate restriction endonucleases and the resulting fragments are electrophoresed on a high-resolution gel; obtained pattern are digitalized and input into a database; and clones with a common pattern are linked to each other in computers. However, this method also requires a vast amount of labor and cost to cleave all the clones, radiolabel them at their ends, and perform autoradiography.

Disclosure of the Invention

The present invention has been made in view of the above-mentioned situation. An objective of the invention is to provide a method for performing electrophoresis of a large number of nucleic acid samples efficiently and inexpensively to detect nucleic acids of interest, and an electrophoresis apparatus used for the method. In the preferred embodiment of the method, the invention provides a means to detect polymorphism in genomic DNAs using the method, a means for genetic analysis, a means for constructing genetic maps, a means for identifying a genomic clone that corresponds to a particular band, and a means to construct a group of organized contigs covering an entire genome.

The present inventors, after conducting extensive studies to solve the above-mentioned problems, thought that the electrophoresis using a small-sized gel, which is generally used in electrophoresis of proteins, could be appropriately used to perform electrophoresis of a large number of nucleic acids efficiently and detect nucleic

acids of interest.

Thus, the present inventors constructed an electrophoresis apparatus for electrophoresis of nucleic acids, where plural 10- to 30-cm square gel plates are installed on the electrophoresis apparatus at a time and 32 or more nucleic acid samples per gel plate are electrophoresed with the electrophoresis apparatus simultaneously. Using this apparatus, the inventors detected nucleic acid markers near a nonpathogenic gene in *Pyricularia oryzae* Cavara or near the brittle culm (kamairazu) mutant gene. As a result, it was found out that polymorphic bands could be detected at remarkably higher efficiency compared to that in the conventional RAPD method.

Linkage analysis on the detected polymorphic bands revealed that, among the detected polymorphic bands, even those at particularly proximal position to the target gene could be obtained by this method at remarkable efficiency compared to conventional methods.

In addition, the method of the present invention was found to be useful in isolation and specific amplification of various important polymorphic bands, such as those detected by the above method. For example, the present inventors used this method, thereby isolating several bands that identify major 10 rice breeds; designed primers according to the sequence information of the band specific to one of the 10 breeds, Akitakomachi; and performed PCR using genomic DNAs obtained from the 10 breeds as templates. As a result, specific PCR-amplification products were found only for Akitakomachi. Thus, it was revealed that the method of the present invention could efficiently provide polymorphic bands, which could be used for easy identification of rice breeds.

Further, the present inventors found out that the method of the present invention could be used to obtain nucleic acid markers for constructing a genetic map of an organism or constructing contigs covering the entire genome of an organism.

Thus, the present invention relates to a method for performing electrophoresis and detection of a large number of nucleic acid samples at high efficiency and low cost, and to an electrophoresis apparatus used for the method and its use. More specifically the present invention provides:

(1) a method for electrophoresis of nucleic acids, said method comprising the following steps:

a) electrophoresing nucleic acid samples using an electrophoresis apparatus on which plural 10- to 30-cm square gel plates are installed at a time and with which 32 or more nucleic acid samples per gel plate are electrophoresed simultaneously, and

b) detecting nucleic acid bands on the gels after the electrophoresing;

(2) the method according to (1), wherein the electrophoresing is performed using gels with discontinuous buffer system;

(3) the method according to (1), wherein the nucleic acid samples are single-stranded DNAs prepared by dissociation of double-stranded DNAs through denaturation and the electrophoresing is performed using denaturing gels;

(4) the method according to (1), wherein the detecting of the nucleic acid bands on the gels is performed by fluorescent staining or silver staining;

(5) the method according to any one of (1), (2), or (4), wherein the method is performed in order to detect a polymorphism of genomic DNAs among test individuals;

(6) the method according to (3), wherein the method is performed in order to detect a polymorphism of genomic DNAs among test individuals;

(7) the method according to (5), wherein the nucleic acid samples are DNA fragments amplified by AFLP method;

(8) the method according to (5), wherein the nucleic acid samples are heteroduplex DNAs;

(9) a method for preparing DNA fragments comprising a polymorphism, said method comprising a step of isolating, from gels, DNA fragments comprising a polymorphism detected by the method according to any one of (5) through (8);

(10) a DNA fragment comprising a polymorphism among test individuals, said DNA fragment being isolated by the method according to (9);

(11) the method according to any one of (1) through (8), wherein the method is performed in order to carry out genetic analysis;

(12) the method according to (11), wherein the genetic analysis

is F2 analysis, RI (recombinant imbred) analysis, or QTL (Quantitative Traits Loci) analysis;

(13) the method according to any one of (1) through (8), which is performed to construct a genetic map of an organism;

(14) a genetic map of an organism, said genetic map being constructed by using, as markers, bands of genomic DNAs comprising a polymorphism detected by the method according to (13);

(15) a method for selecting, from a genomic DNA library, a clone corresponding to a particular nucleic acid band on a gel detected by the method according to any one of (1) through (8), said method comprising the following steps:

a) dividing a genomic DNA library of a particular organism into plural sublibraries each of which has a size of 1 or less genome of the organism;

b) assigning, to all clones included in each of the sublibraries, a row number, a column number, and a plate number of the sublibrary, wherein the row, column, and plate are referred to as X coordinate, Y coordinate, and Z coordinate, respectively;

c) detecting a band by collecting clones representing a particular row of all plates (X-coordinate clone group), clones representing a particular column of all plates (Y-coordinate clone group), and all clones on a particular plate of one sublibrary (Z-coordinate clone group); by extracting DNAs from each of the collected clone groups to obtain coordinate samples; by preparing a genomic DNA from the organism as a control; and by electrophoresing the coordinate samples and the control in a line using the method according to any one of (1) through (4);

d) determining a clone in each of the X-coordinate clone group, the Y-coordinate clone group, and the Z-coordinate clone group, said clone corresponding to a band with the same mobility on the gel as that of the nucleic acid of interest in the control; and

e) selecting, from the sublibrary, a clone corresponding to the determined three-dimensional coordinate;

(16) the method according to (15), wherein the method is performed in order to construct contigs covering the entire genomic DNA of a particular organism; and

(17) an electrophoresis apparatus for electrophoresis of nucleic acids, wherein plural 10- to 30-cm square gel plates are installed on said electrophoresis apparatus at a time and 32 or more nucleic acid samples per gel plate are electrophoresed with said electrophoresis apparatus simultaneously.

1. ELECTROPHORESIS METHOD AND APPARATUS

The electrophoresis apparatus of this invention is that on which small-sized polyacrylamide gel plates (10- to 30-cm square, standard size: 18-cm square) are installed and with which 32 or more (standard: 64) test samples plus several (standard: 2 to 4) size markers per gel are electrophoresed at a time, as well as with which a large number of (standard: 256) samples are electrophoresed by using such 2 or more (standard: 4) gels simultaneously. An example of the electrophoresis apparatus of this invention is shown in Figs 1 through 4. In this apparatus, a 1-mm-thick gel prepared with an 18-cm square glass plate has 66 wells to allow electrophoresis of 64 samples and 2 size standards at a time. The apparatus also allows electrophoresis of 4 gels at a time. These together allow testing of 256 samples in one cycle.

The discontinuous polyacrylamide electrophoresis system (Laemmli , U.K. (1970) Nature 227: 680-685), which is usually used for electrophoresis of proteins, may be used for the gel of this invention. The use of this gel allows high-resolution electrophoresis of as much as 10 µl of a test sample, even in narrow lanes of 1 mm in width. It also increases the sensitivity of band detection.

In the electrophoresis of nucleic acids according to this invention, it is preferable to perform two-layer electrophoresis using concentrated gels (Tris-HCl pH 6.8, 0.5 M) and isolation gels (Tris-HCl pH 8.8, 1.5 M) to improve band resolution of the nucleic acids on the gels. In electrophoresis of nucleic acids, the nucleic acids may remain in double-strand or denatured into single-strand, depending on the purpose. In the latter case, denaturing gels (gels containing 6 to 8.5 M urea) are used in the electrophoresis. When polymorphism in nucleic acids is detected, electrophoresis with heteroduplex allows highly sensitive detection, which can detect minor polymorphisms that

may be undetectable by conventional methods.

In the nucleic-acid electrophoresis of this invention, it is preferable to use silver or fluorescent staining to detect nucleic acid bands after electrophoresis. Silver staining allows highly sensitive detection in a short period of time (1 to 2 hours), requires less expertise and is safer compared to methods using isotopes. In addition, the material may be later dried to provide a preservation sample as well as to increase the sensitivity. With fluorescent staining, results are obtained in approximately 30 minutes of staining. With its high recovery rate of nucleic acid, fluorescent staining is suitable for excision and collection of bands.

In the method of this invention, the efficiency in the step of loading amplified DNA samples and such on the gel may be improved drastically, for example, by designing an electrophoresis gel comb such that 2 or more lanes are loaded at the same interval (9 mm) as those of the 96 (8 x 12) well microplate.

2. DETECTION OF NUCLEIC ACID MARKERS

To detect polymorphism in nucleic acids using the method of this invention, combination with AFLP (Amplified Fragment Length Polymorphism: Vos P, et al. (1995) Nucleic Acids Research 23: 4407-4414) can conceivably yield the highest efficiency although combinations with other forms of detecting nucleic acid polymorphism are also possible.

The use of AFLP, for example, yields approximately 50 amplification bands per lane from a genome of a size on the order of that of rice (450 MB), with primers to amplify genome fragments having a 6-base cleavage site at one end and a 4-base cleavage site at the other end, when 3-base selective nucleotides are used at each end (refer to the below-mentioned formula). A system comprising a standard 4-gel 256-sample lane would yield approximately 12,800 bands in one electrophoresis cycle (Example 2).

Entire genome size ÷ the number of cleavage sections by 6-base restriction endonuclease ÷ selection rate at both ends of genome fragment by 3 nucleotides = the number of bands

$$(450\ \text{MB} \div 4^6 \times 2 \div 4^{3 \times 2} \fallingdotseq 50)$$

This number compares to the amount of information obtained from several gels from RLGS method. In addition, in the method of this invention lanes to be compared may be placed next to each other, permitting easy, direct comparison of raw data without the need for special reading means such as a special reading apparatus.

As shown in Fig. 6 (Example 2), in actual AFLP experiments where rice genome DNAs were cleaved by EcoRI and MseI, the number of bands obtained was close to that estimated with the above formula. However, band with certain clarity were about half of them.

Performing PCR with 5 to 10 µl of sample using a 0.2-ml microplate for 96 samples saves costs of nucleotides and such enzymes as heat-resistance DNA polymerase and increases efficiency by allowing transfer of 8 samples to the gel at a time using an 8-channel 10-µl pipette. Thus, one electrophoresis cycle with 4 gels may be easily performed in one day.

After electrophoresis is completed, the 4 gels may be silver-stained simultaneously in one container, which is an efficient, low-cost and simple procedure. In general, commercially available silver-staining kits for protein may be used.

As nucleotide primers for AFLP, any nucleotides corresponding to selected restriction endonucleases may be used. Further, any selective sequence with a length of 1 to several bases may be added to the 3' end. This allows almost infinite combination of primers. For a plant with a genome of 1 GB or less size, by using EcoRI and MseI as a 6-base restriction endonuclease and a 4-base restriction endonuclease, respectively, and by using PCR primers having 3-nucleotide selective sequences, each of which has 64 combinations, the amplification would be performed in 64 x 64 = 4,096 combinations. This allows search for 10,000 to 20,000 polymorphic bands in genetic analysis between parents with a polymorphism rate of 5 to 10%. This number is practically sufficient for physical map construction and exceeds the numbers of makers on conventional genetic maps by an order of magnitude or more.

Important bands such as those near the target gene, obtained from bulk analysis (Michelmore RW, et al. (1991) Proc. Natl. Acad. Sci. USA 88: 9828-9832) or other method, may be cut out after staining

for further analysis: The cut-out gel is crushed and provided for extraction in appropriate buffer for another PCR procedure; the PCR product is then inserted into appropriate plasmid, conjugated, and thus isolated for sequencing (Example 4).

With combinations described above, it is possible to perform efficient electrophoresis as follows, for example: A set of 4 samples, 2 from breeding parents and 2 from mixtures of approximately 10 individual of F2 homogenous individuals, dominant and recessive each, are prepared; electrophoresis is performed with 4 lanes / primer-pair; with a standard electrophoresis apparatus (4 gels, 256 lanes), electrophoresis with 64 primer pairs may be performed at a time. In this example, tests with all 4,096 combinations of selective primers are completed in 64 times of electrophoresis (total approximately 2 month). In a breeding combination with approximately 10% polymorphism, such as indica-japonica crossing, this corresponds to scanning and searching for 20,000 polymorphic bands in the entire genome. This efficiency exceeds those of conventional methods by 1 to 2 orders of magnitude and permits gene isolation in a short period of time, regardless of presence or absence of a genetic map. The cost for search of all 4,096 combinations is also low, being approximately ¥400,000.

3. GENETIC ANALYSIS

1) F2 analysis

To determine a genetic distance (or distance on the map) between a combination of given genes or polymorphism markers, F2 analysis is most commonly used. The use of F2 analysis in accordance with this invention even more efficiently determines a distance between 2 polymorphism markers or between a polymorphism marker and a gene.

Like in general F2 analysis, a line (line A) that contains a given gene and another line (line B) that does not contain the gene or has distinct differences in traits related to that gene and that has appropriately small differences in traits from those of A are selected and crossed to produce a large number of F2. The number of F2 individuals to be analysed depends on the precision of the analysis. To achieve 1 CM precision, 50 homogenous individuals (usually recessive

homogenous individuals) are sufficient to determine the genetic distance between a gene of interest and a polymorphism marker, by determining the recombination frequency between the gene and the marker on 100 chromosomes. In an analysis of recessive homogenous individuals, if no recombination occurs, all of the individuals should have the same marker as of the parent with a recessive trait; with the recombination frequency of 1/100, recombination should be observed only in 1 chromosome of 1 individual.

Since the standard system of this invention is capable of analyzing 256 individuals at a time, if such a number of homogenous individuals are available, recombination of polymorphism markers may be tested on 512 chromosomes in one electrophoresis cycle, which yields 0.2 CM precision. The use of AFLP for polymorphism markers allows testing with very small amount of sample DNAs (100 ng or less); genomic DNAs are prepared from each individual and double-digested with EcoRI and MseI enzymes. Adapters matching with each of the enzymes are then coupled with the genomic DNA fragments. Using first primers matching these adapters, the first PCR is performed to amplify the genome fragments. Then second primers are prepared by adding a selected 1- to 4-base sequence to the 3' end of each of the first primers. Using these second primers, only a part of the genome fragments which corresponds to the selected sequences are amplified by PCR. The PCR products are then separated by molecular weight with electrophoresis of this invention. After electrophoresis is completed, the gel is stained by such as silver staining and examined for recombination in the polymorphism marker of the interest.

In accordance with this invention, F2 analysis with small amount of DNAs from as much as 256 individuals is completed in one electrophoresis cycle with a precision of 0.2 CM level. Further, the process does not require blotting or autoradiography. In addition, after staining and drying, the 4 gels may be cabinet size and filed in albums, which permits easy analysis of the results.

2) RI analysis

An RI (Recombinant Inbred) line refers to that obtained by self fertilization of F2 individuals for several generations which made

by breeding of different lines. As a result of repeated self-fertilization, most of loci are homozygous. Because of a limited number of heterozygotes, the separation ratio of dominant-recessive traits is 1:1. Thus, genes of interest are homozygous even in dominant individuals, which makes them applicable to the gene analysis of this invention. This gene analysis is more precise than F2 analysis, where the proportion of heterozygous individuals is 1/2 and the separation ratio of dominant-recessive traits is 3:1. When an appropriate RI line is available, gene analysis using genomic DNAs extracted from a number of the RI line (RI analysis) may be performed in accordance with this invention in the same manner as in F2 analysis.

3) Narrowing-down of proximal markers

To narrow down, by F2 analysis, proximal markers to the gene of interest obtained in bulk analysis, and ultimately map the most proximal markers, electrophoresis of the present invention is useful; the ability to perform electrophoresis with 256 samples at a time in the standard method provides very high efficiency in the gene analysis. With rice (in case of indica-japonica crossing), as described in 2., given that 20,000 bands of polymorphism markers evenly distribute over the entire 2,000 CM genome, the number of markers present in the 20 CM region, 10 CM on each side of the gene obtained from bulk analysis, is estimated to be 200 bands. To narrow down such a number of proximal markers to the most proximal markers, 14 recessive homogenous individuals of F2 generation per marker and DNAs from each parent are run in 16 lanes as a first step to examine for presence or absence of recombination between the target gene and the candidate marker. Thus, recombination frequency for 28 chromosomes per maker is determined. The resolution in this process is approximately 3.5 CM. Thus, 4 bands of candidate markers are tested per gel, which corresponds to testing of 16 bands of markers per electrophoresis cycle. Simultaneously, the individuals having recombination at the most proximal site to the target gene is identified. After this process, using merely 8 lanes for 6 individuals with proximal recombination and for the parents, testing of the rest 192 markers is completed in 6 cycles of electrophoresis.

Given that positions of markers and chromosome recombination distribute evenly, it is expected that, as a result of above-mentioned screening, approximately 40 bands of markers be obtained from an 8 CM region, 4 CM on each side of the target gene. The second step is to map each of the markers remaining from the first screening at 1 CM precision. First, using one gel each, 12 markers are run in 3 cycles of electrophoresis in order to obtain AFLP products of 62 recessive homogenous individuals of F2 generation and the parents. This produces a fine map for the 8 CM region around the target gene, which reveals individuals having recombination in the region immediately close to the gene, approximately within 1 CM from the gene. Using such individuals with proximal recombination, the rest 28 bands are analyzed to identify their locations on the map. One electrophoresis cycle of 224 lanes, which is 8 lanes (6 individuals with the most proximal recombination plus parents) times 28 combination, completes the analysis of the region around the target gene.

Thus, 200 candidate proximal markers selected from 20,000 polymorphic bands by bulk analysis may be narrowed down to most proximal markers within approximately 1 CM from the target gene by approximately 11 cycles of electrophoresis.

In gene isolation, if the genome size of the organism of interest is 1 GB or smaller, the average 1 CM in such markers should be 500 kB or less. Therefore, clones proximal to the target gene may be selected from a genome library of BACs (bacterial artificial chromosomes) which have approximate average insert size of 150 kB, to construct a contig.

It is also possible to increase the analysis precision to 0.2 CM level by using 256 recessive homogenous individuals. In this case, by selecting about 4 bands of appropriate proximal markers and performing 4 electrophoreses cycles, individuals having recombination most proximal to the target gene is identified. Using such individuals with proximal recombination to check recombination in the presumed bands most proximal to the gene in the same manner as of the second step, the most proximal markers are identified easily by 1 electrophoresis cycle.

Herein the principle of the invention has been described with

two assumptions; bands obtained by electrophoresis distribute at almost even frequency over the genome or chromosome, and recombination in chromosome occurs uniformly over the genome. Actually, however, distribution of bands and recombination sites in a chromosome is not uniform. Thus, the distance between a proximal marker and the target gene, where the distance is obtained for the number of F2 individuals used, tends to be larger than that expected for uniform distribution.

4) Application to marker breeding

A proximal marker for a particular trait gene obtained in the above described manner may also be used as a highly reliable marker in conventional breeding by mating. Also, when marker analysis is performed on a large number of mating offspring, the analysis method using genome scanning allows easy analysis of a large number of individuals with small amounts of DNA samples. This would result in substantial increase in operation efficiency and decrease in cost compared to conventional techniques using the RFLP method.

5) Application to QTL (Quantitative Traits Loci) Analysis

A QTL is a locus of quantitative trait, where expression of the trait is not as strong as to be qualitative and, in most cases, several loci are involved in expression of the trait. QTL analysis is performed to analyze which loci at which location on the chromosome how much contribute to expression of the trait.

QTL analysis requires that polymorphisms in a large number of individual offspring of mating should be tested for a large number of nucleic acid markers distributed evenly over the entire genome. For example, to perform QTL analysis using markers distributing over a 2,000 CM entire genome at the marker density of approximately 10 CM/marker, 200 markers need to be tested with at least about 50 F2 individuals. Performing this analysis with RFLP markers would require 200 times of hybridization with membranes blotted with nucleic acids of 50 individuals. Since one cycle of hybridization requires 2 days, even if 4 membranes were processed at a time, the whole process would require 100 days. Even if a membrane could be repeatedly used for 10 times, it would be necessary to collect as much as 100 μg of nucleic

acids from each of 50 individuals to prepare 20 membranes. This would require enormous labor.

By using AFLP method in the present invention, a few polymorphic bands are obtained in a lane when an organism with approximately 10% of polymorphism, such as indica-japonica crosses of rice, is studied. Therefore, by using appropriate primer pairs, search for 200 markers is completed by electrophoresis with a little over 50 pairs of primers. Since 5 primer pairs may be used in 1 electrophoresis cycle (50 x 5 = 250 lanes), 10 cycle (total 10 days) of electrophoresis would complete examination of all the markers for 50 individuals. In addition, 1 μg of DNAs per individual is enough to perform this process.

Specifically, a genome map for the organism with AFLP markers must be prepared prior to performing this method. In case where such a map is not available, a map may be readily constructed according to this invention as described in 5 below. Markers are selected from the map at a desired density. In this case, it is efficient that as small a number of primer pairs as possible are selected so as to cover the entire genome.

To achieve an appropriate polymorphic frequency in breeding, it is recommended to select two genetically distant lines one of which strongly expresses the quantitative trait of interest and the other shows very little expression the trait. Breeding of too distant lines may inhibit smooth isolation of the markers. Using approximately 50 individuals (the number varies depending on purpose of the analysis) of the F2 or RI line obtained from this breeding, the trait of interest is quantitatively analyzed. Also, using genomic DNAs prepared from each individual, the genomic fragments are amplified by AFLP method as described in 3.1), and are subjected to electrophoresis using the system of this invention for examination and recording of polymorphisms in the marker bands.

To indicate contribution of each locus near each marker to the trait of interest, the number of traits of interest is multiplied by the number of polymorphisms in one of the parent in each band, and the product is plotted for each of the marker bands on the map.

4. APPLICATION TO IDENTIFICATION OF BREEDS AND/OR LINES

This invention may be used to efficiently search for necessary markers to identify a particular breed and/or line of an agricultural or livestock product or various organisms. AFLP is performed with DNAs obtained from a large number of breeds to be compared and electrophoresis is performed simultaneously using the electrophoresis apparatus of this invention to compare dozens of bands for dozens to 100 or more breeds at a time. Thus, even with a large number of breeds, specific recognition bands are easily selected.

In this application, although combination of n bands enables identification of up to 2n breeds and/or lines theoretically, it may be slightly less than that practically.

When the breeds and/or lines to be compared are so close to each other that obtaining polymorphisms by conventional methods such as AFLP is difficult, products of AFLP, RAPD, or any other method from genome of the compared sample may be mixed, heated, and then cooled to provide heteroduplex; the heteroduplex may be subjected to electrophoresis for comparison with the bands of interest. Thus, differences between breeds and/or lines are detected sensitively, efficiently, and simply.

When a particular band with high identification ability is identified, the band may be isolated in a manner described in 7 so that the band may be PCR-amplified with certain primers. After the PCR amplification, simple agarose gel electrophoresis enables identification of the breed within approximately 1 hour. Further, the primers may be labeled with certain fluorescent labels so that presence or absence of the band amplified from the primers is determined by a PCR apparatus equipped with a fluorescent detector. In this method, 2 to 30 minutes of PCR would identify the breed and/or line without the need of electrophoresis.

5. CONSTRUCTION OF A GENETIC MAP OF AN ORGANISM

To construct a genetic map (more accurately, nucleic acid marker map) covering the entire genome of an organism, it is necessary, like in QTL analysis, to analyze several hundreds to thousand or more markers with F2 individuals, in which the number of F2 individuals depends on the resolution of the required map. In most higher plants,

especially in major crops, the total length of genome is between 1,500 and 2,000 CM. Thus, to construct a genomic map in the density and precision of about 1 CM with conventional RFLP method, it would be necessary to prepare membranes from 100 or more F2 individuals and repeatedly blot the membranes for about 1,800 markers. Since 1 cycle of blotting, including preparation of the probes, requires 4 days, the whole process would require 1,800 days even if 4 probes were processed at a time. Also, a very large amount of nucleic acids is needed for this process; 1 mg or more DNAs is needed for each F2 individual. This is one of the reasons that genetic map construction has been conducted by a team comprising dozens of persons over several years.

According to this invention, with breeding parents with approximately 10% polymorphism, for example, a pair of primers indicates 4 to 5 polymorphisms on average. Hence, by conducting a pre-examination to select primer pairs that indicate 6 or more polymorphisms and by mapping of 128 F2 individuals with the standard model using the selected primer pairs, 12 or more markers are mapped in 1 electrophoresis cycle with 2 primer pairs. Thus, a fine map including 1,800 markers would be completed in total 150 days with 1 person, which is 10 times or more efficient than in conventional methods. A map containing approximately 300 markers would be completed in approximately 1 month with one person.

Specific procedure is as follows: Two lines at an appropriate genetic distance are bred to obtain a RI line or F2 individuals to the number according to the precision of the required map. To minimize the potential of hybrid sterility and distortion in separation ratio, the average polymorphism rate between the two lines should be up to 10% or slightly higher. To achieve the precision of approximately 0.5 to 1 CM, 64 to 128 individuals are generally enough. Genomic DNAs are prepared from these individuals and amplified by the AFLP method in the same manner as already described for F2 analysis and RI analysis. The amplified PCR fragments are provided for electrophoresis, staining, and analysis by the system of this invention. This is done as follows: The prepared genomic DNAs are stored in 96-well microplates. A part of the DNAs (approximately 0.1 μg) are double-digested by EcoRI and MseI. Adapters are added to the cleavage ends, and pre-amplification

is performed using primers with the same 5' ends as of the adapters. Then, using the PCR products of the pre-amplification as templates, secondary amplification is performed using selective primers with an appropriate number of selective bases. The amplified products are applied for the electrophoresis apparatus of this invention. The use of 8- to 12-channel pipettes or 8- to 12-channel microsyringes increases efficiency of gel loading. When 64 samples are tested, approximately 20 markers may be processed with 4 primer sets in 1 electrophoresis cycle; when 128 samples are tested, 10 markers with 2 primer sets. Genetic map construction software such as MAPL or MAPMAKER may be used for polymorphic band analysis to improve efficiency.

A reference case using a former system of the present invention is shown herein (Fig. 11). Eight 15-lane gels were used in this system. Using 99 RI lines from breeding of two barleys, Azumamugi and Kanto Nakate Gold, an AFLP map with 227 markers was constructed in approximately 2 months. This AFLP map was integrated with an already-existing map including 45 markers to produce a map with 272 markers. In this case, the AFLP map was completed in approximately 2 months by 1 person, which is approximately 40 times higher efficient than conventional map constructing methods, such as those using STS markers. The method of this invention is even more efficient than this former system.

6. APPLICATION TO ORGANISMS WITH LARGER GENOMIC SIZES

The present invention conceivably exerts its potential most when combined with AFLP. A future of AFLP is in that genome is double-digested into fragments by EcoRI (which recognizes 6 bases) and MseI (which recognizes 4 bases); adapters matching the cleaved sites are added to the fragments; the adapter sequences are used as PCR primers in the first amplification so that all the fragments are amplified; selective primers in which 1- to several-base selective sequence are attached to the 3' end of the first primers are used as second primers in the second amplification; and thus only genome fragments having sequences that match the selective primers on both ends are specifically amplified. In case of an organism with a relatively small genome size (such as rice: 450 MB), selective primers

in which 3-base selective sequences are attached to the EcoRI and MseI sites on both ends are used as the second primers. Thus, as described in "2. Detection of nucleic acid markers", approximately 50 bands on average are selectively amplified. With genome of a size similar to that of filamentous fungi, such as *P. oryzae* Cavara (40 MB), applying only 1 base to either of the cleaved sites would produce 50 x (40MB/450MB) x 16 ≒ 70 bands. To organisms with larger genome sizes, this method may basically be applied simply by increasing the number of bases of the selective sequence for AFLP primers to 3 or 4.

Alternatively, conditions of electrophoresis may be changed. Usually, in electrophoresis of PCR products, it is convenient to perform electrophoresis of PCR products under non-denaturing conditions leaving the material double-stranded. However, with an organism with a larger genome size (such as barley: 5.5 GB), bands may not be clearly separated if the PCR products are double-stranded. Therefore, by using denaturing gels containing 8.5 M urea and placing the DNA sample in 90°C for 3 minutes in presence of 50% formamide so that the DNAs are separated into single strands, and then subjected to electrophoresis a large number of bands are clearly distinguished (Example 4).

In practice, as genome size becomes larger, difficulty of AFLP increases. Thus, it is recommendable to combine the above two strategies.

7. ISOLATION AND SEQUENCING OF IMPORTANT BANDS FOR PREPARATION OF SCAR MARKERS

In the embodiment of this invention such as those described in above sections 3., 4., and 5., bands with information particularly important in each case are identified: bands proximal to a particular gene (in section 3.); bands applicable for breed identification (in section 4.), and band that serve as landmarks on a genetic map (in section 5.).

These bands may be cut out from the gel to provide SCAR (Sequence Characterized Amplified Region) markers as follows: DNA material in the cut-out gel is eluted by crush and extraction, freeze and thaw, electrophoresis, or other method and provided for PCR amplification

using the same primer pair. The PCR products are inserted into appropriate cloning vectors for sequencing. After the sequence is determined, the both ends are used as primers of SCAR markers (Example 3). In this case, the use of a fluorescent dye such as cyber green facilitates extraction of nucleic acids. Compared with AFLP method using sequencer, in which bands are identified but cannot be removed, the method of this invention is more simple while fully utilizing the important features of AFLP.

8. ISOLATION OF GENOMIC CLONES CONTAINING SPECIFIC BANDS

When bands containing important information are found, such as illustrated in above section 3., 4., and 5., the genomic clones containing the bands may need to be isolated. Especially, to isolate a gene with important functions by positional cloning, it is necessary to prepare a series of clone contigs as follows: The most proximal markers are identified in both sides of the gene by methods such as 2 or 3; using these markers, a clone containing the bands are selected from a genomic library such as the BAC library; using the markers at both ends of the clone, next contiguous clone is identified, and this process is repeated (walking); then a series of clone contigs that connect between markers present on both sides of the gene (flanking markers) is prepared.

When high-density membranes in which component clones of a genomic library are organized and linked by colony hybridization have been prepared, the target clone may be picked up by performing hybridization on the membranes using amplified bands as probes. These bands may be prepared by direct PCR as described in section 7., or may be amplified as SCAR markers.

Alternatively, clones containing the band of interest may be identified as follows, without the need for hybridization: According to section 9., the genomic library of the target organism is divided into sub genomic libraries; coordinate markers are prepared by mixing clone DNAs of the sub genomic libraries such that the clones correspond to row, column and plate numbers of the microplates; these coordinate makers are provided for AFLP with same primer pair using genomic DNAs as control and then for electrophoresis of the present invention;

thus clones containing the band of interest is identified according to the coordinates of raw, column, and plate in which the same band as of the control is amplified.

9. PREPARATION OF CONTIGS COVERING THE ENTIRE GENOME OF AN ORGANISM

If contiguity status of all of these component clones is unclear in accordance with a genomic library of an organism, such library as those based on BAC, contigs are linked together to construct a larger contig (physical map) covering the entire genome. A physical map constructed in this manner represents a reproduction of the genomic structure of the organism. Completion of such a map makes it much easier to isolate and to identify important functioning genes and facilitates handling of the genome of the organism, even before the entire genome is sequenced. In fact, once the entire genome contig is completed, following genome sequencing may be almost completely mechanized. Thus far, construction of an entire genome contig has required a large research group consisting of dozens of persons. With the present invention, however, a contig covering the entire genome of an organism whose genome size is close to that of rice (approximately 500 MB) may be constructed in about a year with one person.

Construction of a whole genome contig has been difficult mainly because it has been difficult to obtain specific labels to mark each clone composing the genomic library. In this invention, by combining with AFLP, as many as 30 to 50 highly specific bands per line are obtained in one cycle, which bands are labeled with two parameters, 3-base sequence on both ends and the number of bases (length). Each of these specific bands may be associated with a component clone of the genomic library and the distribution density of the bands may be made sufficiently smaller than the length of the BAC clone. Thus, it is not difficult to link clones having common markers together.

To achieve almost one-to-one correspondence between the AFLP bands and the library clones, the genomic library for the entire genome, which generally has several genome equivalents or more, is divided into several sublibraries of approximately 1 genome equivalent each. Each component clone of the sublibraries is uniquely identified by coordinate of row, column, and plate numbers of the microplate.

Therefore, using row, column, and plate numbers as coordinate axes, small amounts of DNAs are collected from groups of clones with a common axis coordinates, and mixed to provide coordinate samples representing positions on each axis. By performing genome scanning of the present invention using these coordinate samples as templates and comparing AFLP patterns with control lanes, which are prepared by using whole genome DNA as templates, placed on both sides of the coordinate lanes, clones corresponding to specific bands from the genomic DNA templates are readily detected from the sublibrary. When there are 2 to n corresponding clones in the sub-group, 2^3= 8 to n^3 clones correspond to the bands. In this case, by removing these 8 clones and performing second electrophoresis of the genome scanning method, the truly corresponding 2 clones are identified. Specific procedures to prepare coordination samples of a sublibrary are shown in Example 5.

With a library of a genomic size on the order of that of *P. oryzae* Cavara (approximately 40 MB) and of the average insert size of 120 kB, a sublibrary of approximately 1 genome equivalent would contain approximately 300 clones, which may be stored in 8 rows, 6 columns, 6 half-plates. Thus, if one gel contains 22 lanes, 20 for coordinate samples and 2 for controls, 2 gels are sufficient to perform electrophoresis of sublibraries of 6 genome equivalents, i.e., the whole genome library. In other words, if 1 electrophoresis cycle identifies and coordinates approximately 60 bands, with 2 primer pairs, 1500 bands are processed in 25 electrophoresis cycles. This corresponds to 40 MB / 1500 bands = 27 KB/band density, which is equivalent to identifying 4.4 bands in average in a BAC clone of an average size of 120 kB. This is expected to sufficiently cover the whole genome, given that there is clone redundancy of average 6 genome equivalents (refer to Example 5).

When the genome size is on the order of that of rice (450 MB), if the library is 6 genome equivalents with an average insert size 150 kB, electrophoresis of 6 sublibraries is completed in 1 cycle by preparing a matrix of 16 rows, 12 columns, 16 x 2 plates using 32 plates, which corresponds to approximately 1 genome equivalent. This allows identifying about 25 bands per cycle. Consequently, 200 electrophoresis cycles would identify 5000 bands, which corresponds

to 450 MB / 5000 bands = 90 kB/band density. This density should be sufficient to construct long contigs from a library in which average 150 kB clones are present at 6 times redundancy.

Brief Description of the Drawings

Fig. 1 is an elevation view and a top view of the main body of a standard electrophoresis apparatus used in genome scanning, capable of processing four 18 cm square gels at a time. On each gel, wells for sample application are made with a comb which can load 66 to 68 samples in 1-mm width. Since 2 to 4 lanes are used for molecule weight markers or such, 64 samples/gel, i.e. 256 samples on 4 gels are substantially processed at a time. PCR-amplified samples may be efficiently applied on the gels by the use of 8-channel micropipettes.

Fig. 2 is a drawing of gel end parts of the standard electrophoresis apparatus used in genome scanning.

Fig. 3 is a drawing of the comb of the standard electrophoresis apparatus used in genome scanning.

Fig. 4 is a photograph of a completed standard electrophoresis apparatus used in genome scanning.

Fig. 5 is an electrophoratogram showing the primary screening using genome screening for candidates of proximal markers to avrPib, a nonpathogenic gene of *P. oryzae* Cavara, obtained through bulk analysis. Genome scanning combined with AFLP was performed on 4 primer pairs, A, B, C, and D, which have different proximal marker candidates, using, from the left, parent lines of avrPib - and +, bulk of avrPib - and +, and each six F1 lines of avrPib - and +, respectively. The triangle marks in the figure indicates bands for marker candidates. The primary screening of the primer pairs A, B, and C indicated coisolation with the genotype as expected. On the contrary, recombination was observed in one - line and one + line with the primer pair D (solid triangles). Note that each primer pair produced 5 to 60 bands. Three to 4,000 bands may be scanned on one gel. Since the apparatus is capable of processing 4 gels at a time, 12 to 16,000 bands are scanned in one cycle. The numbers of selective nucleotides used on these primer pairs were 3 on the EcoRI side and 1 on the MspI side. The polymorphism rate between the breeding lines used herein was approximately 5%.

Since the whole genome size of *P. oryzae* Cavara is approximately 500 CM, about 500 polymorphic bands are obtained with 256 primer pairs, i.e., the whole genome is covered with the marker density of 1 CM / band.

Fig. 6 is a fine map around avrPib, a nonpathogenic gene of *P. oryzae* Cavara, where the map was obtained by mapping proximal markers to the gene by RAPD method and genome scanning using 125 F1 lines. As shown in Table 1, proximal markers were searched by RAPD method using 700 primers and by genome scanning using PCR with 251 primer pairs. The RAPD method required 3 months for the search and mapping while genome scanning was completed in 1 month. Since only very-clear bands were counted, the numbers of the counted bands tended to be small. An: markers obtained by genome scanning, (Rn) : markers obtained by RAPD method.

Fig. 7 is an electrophoretogram showing an example of bulk analysis for search for proximal markers to bc-3, a gene of Kamairazu, mutant for cellulose synthesis in rice. For each primer four lanes are shown: from the left, mutant parent line (M11: japonica), mutant (recessive) homogenous bulk, wild-type (dominant) homogenous bulk, and wild-type parent line (Kasalath: indica). The arrows in the figure indicate candidates for proximal markers, which show distinctive difference between the bulks. There are 20 to 30 bands per lane when only distinctive bands are counted; when narrower bands are included, average 50 bands are seen per lane, which reflects the theoretical value.

Fig. 8 is a photograph of genome scanning performed to search for genes responsible for the two-rowed spike trait in barley. As backcrossing hybrids of Azumamugi (six-rowed) and Kanto Nakate Gold (two-rowed) with Azumamugi (six-rowed) over 7 generations, lines with two-rowed spike were selected to establish a quasihomogenous genetic line having the two-rowed spike trait with background of Azumamugi. This line and the backcross parent Azumamugi are compared by genome scanning in the photograph of Fig. 8. Differences from 16 primer pairs were searched for in 32 lanes, and a very limited number of bands showed differences. The bands with difference are candidates for polymorphic bands in the regions strongly associated with the two-rowed

spike genes.

Fig. 9A shows examples of SCAR (Sequence Characterized Amplified Region) markers obtained by isolating bands that are specific to a rice breed, Akitakomachi. The reverse portions are the prepared primers. Fig. 9B is an electrophoratogram showing bands amplified from nucleic acids of major 10 rice breeds by PCR using the primers shown in Fig. 9A. A specific band was amplified only from Akitakomachi. In "Control 1", a specific band isolated for sequencing was used as a template.

Fig. 10 shows how a clone corresponding to a particular band is selected directly from a genomic library using genomic scanning. In this case, the library of *P. oryzae* Cavara (genome size: 40 MB), which is 6 genome equivalent with an average insert size 120 kB, is divided in to 6 sublibraries of approximately 1 genome equivalent each, and clones corresponding to genomic-scan bands are identified in each sublibrary.

In Fig. 10A, the clone indicated as a dot is expressed as row C, column 4, and plate a. The coordinate sample representing the coordinate of row C is prepared by collecting 10 ng each DNAs from clones in all columns in row C on all half plates. In Fig. 10B, one sublibrary consists of 22 lanes, which include reference genome lanes and coordinate samples for 8 rows, 6 columns, and 6 half plates. Six genome equivalents are loaded on 2 gels and are processed in 1 electrophoresis cycle. The corresponding clone is identified according to the coordinates which show the same bands as of the control.

Fig. 11 is an example of a gene map of barley constructed by 15-lane gel electrophoresis, from which the genomic scanning was developed. Among total 272 markers, 227 AFLP markers were mapped by this method.

Best Mode for Carrying out the Invention

The following Examples are given to further illustrate this invention; however, the present invention is not intended to be limited to the specific Examples.

[Example 1] Fine mapping of nucleic acid markers near avrPib, a

nonpathogenic gene for rice

The nonpathogenic gene in *P. oryzae* Cavara, avrPib, which corresponds to the resistant gene in rice, Pi-b, was searched for nucleic acid markers around the gene and distances between the markers and the gene were determined. The gene avrPib is a nonpathogenic gene in *P. oryzae* Cavara, which corresponds to the resistant gene in rice, Pi-b; resistance of rice to pathogenic *P. oryzae* Cavara depends on recognition of gene products of avrPib by the Pi-b gene in the rice. Hence, mutation in the avrPib gene disrupts resistance by the Pi-b gene. Thus, to investigate the cause of mutation as well as interactions between resistant gene products and nonpathogenic gene products, it is necessary to isolate the avrPib gene from resistance-disrupted lines. By crossing with *P. oryzae* Cavara which were isolated from regions where disruption of resistance was observed and were affecting Pi-b, fine mapping of the nonpathogenic gene was compared between the method of this invention and RAPD method, a representative conventional method.

Lines retaining avrPib (and thus not capable of affecting Pi-b-containing rice) and lines without avrPib (and thus affects Pi-b-containing rice) were crossed to produce a large number of F1 generation lines. These F1 lines were injected to rice having Pi-b to determine presence or absence of avrPib in each line. Then genomic DNA bulks were prepared from avrPib-present group (+) and avrPib-absent group (-), using about a dozen lines for each group. Using 4 lines, i.e., 2 from each parent line and 2 from (+) and (-) bulks, AFLP analysis was performed with 64 x 4 = 256 primer pairs. Since the genome size of *P. oryzae* Cavara is approximately 40 MB, which is about 1/10 of that of rice, the number of primer pairs required to cover the entire genome is about 1/16 of the case in rice.

For comparison, approximately 540 primers were used for RAPD (Random Amplified Polymorphic DNAs) amplification. In RAPD method, 144 lanes, for example, may be processed at a time using a large-sized submarine gel while the bulk method requires 540 x 4 = 2160 lanes. Hence only stable and distinctive bands were chosen for comparison. Thus, chosen 1860 bands were compared between the bulks over 15 days, which resulted in approximately 100 polymorphic bands between the

parents.

In contrast, the method of present invention compared and searched for approximately 5700 bands in 4 days, resulting in 304 polymorphic bands (Table 1). This indicates that the efficiency of this method is 11 times higher than that of RAPD method (5700/1860 x 15/4).

Table 1

	Used primers (pairs)	Total bands obtained	Polymorphic bands		Average bands / primer (pair)	Percentage of polymorphism (%)
			Total	Linked to *avrPib* (Closely linked)		
RAPD	**539**	**1861**	**101**	**10 (4)**	**3.5**	**5.4**
AFLP	**251**	**5710**	**304**	**86 (41)**	**22.7**	**5.3**

Among these polymorphic bands, 6 bands from RAPD and 41 from AFLP were considered as candidates for proximal bands to the gene, in which the bulk method revealed distinctive difference between the dominant and recessive homogenous groups. These candidate bands were at fist provided for primary screening, using 12 F1 lines as the reference (Fig. 5). The number of F1 lines was increased, and finally, with 125 F1 lines, a fine map around the avrPib gene was constructed (Fig. 6). The RAPD method revealed 2 bands, and the method of present invention 12 bands, within 20 CM from the target avrPib gene. While the band closest to the gene detected by RAPD was at 5.3 CM from the gene, the method of present invention detected a band at 1.8 CM, which was much closer. This distance is considered small enough for construction of a physical map since 1 CM in *P. oryzae* Cavara corresponds to approximately 80 kB.

[Example 2] Fine mapping of bc-3, a Kamairazu (brittle culm) mutant gene in rice

The bc-3 gene is a causal gene for the mutation with which cellulose synthesis required for the secondary thickening of cell walls in the culm of rice is inhibited (Kamairazu or brittle culm). Nucleic acid

markers near the bc-3 gene were searched for by genomic scanning. A japonica bc-3 mutation M11 and an indica wild-type Kasalath were used as crossing parents for F2 analysis. Ten mutant homogenous individuals and eight wild-type homogenous individuals were identified by analysis up to the F3 generation, and their genomic DNA mixtures as well as those of the parent lines were provided for genomic scanning bulk analysis in combination with AFLP. The nucleic acids were double-digested by two enzymes, i.e., EcoRI (which recognized 6 bases) and MseI (which recognized 6 bases). Then, genomic scanning by bulk analysis was performed with 1430 combination of primers having 3-base selective nucleotides contiguously to the ends cleaved by the enzymes. An example of the result is shown in Fig. 7.

As a result, 97 candidates for proximal markers were obtained, among which 50 were usable to detect recombinant markers in recessive homogenous individuals. To narrow down the candidates for markers, each candidate was provided for F2 analysis using 10 recessive homogenous individuals (20 chromosomes) each. Among them, 24 marker candidates were analyzed with 32 recessive homogenous individuals, i.e., 64 chromosomes, and 1 marker was found to be coisolating with (at 0 CM distance from) the target gene (Fig. 7).

[Example 3]

When this invention is to be applied by AFLP to an organism with a very large genome size, such as barley (5.5 GB), if the lengths of selective primers on the both ends of genome fragments are 3 bases each, distinctive bands may not be obtained if the DNAs to be tested remains double-stranded under nondenaturing conditions, which are used for smaller genome sizes, such as that of rice (450 MB). In this case, one approach is to increase the number of selective primers. Alternatively, however, a sufficient band resolution may be achieved even with 3-base selective markers by denaturing DNAs into single strands upon electrophoresis. To provide such denaturing conditions, 6 to 8.5 M urea is added to the gel, 50% formaldehyde is added to the sample buffer, and the samples are placed in 90°C for 3 minutes immediately before electrophoresis.

Fig. 8 is a result of genome scanning performed to search for

genes responsible for the two-rowed spike trait in barley. As backcrossing hybrids of Azumamugi (six-rowed) and Kanto Nakate Gold (two-rowed) with Azumamugi (six-rowed) over 7 generations, lines with two-rowed spike were selected to establish a quasihomogenous genetic line having the two-rowed spike trait with background of Azumamugi. This line and the backcross parent Azumamugi are compared by genome scanning in the photograph of Fig. 8. Differences from 16 primer pairs were searched for in 32 lanes, and a very limited number of bands showed differences. The bands with difference are candidates for polymorphic bands in the regions strongly associated with the two-rowed spike genes.

With barley, average approximately 58 bands are identified under these conditions. Given that approximately 88 bands are obtained in average by AFLP of barley using a large-scale sequencer gel (Castilioni et al. 1998 Genetics 149: 2039-2056), approximately 67% of the bands are identified by this method. The method of present invention is capable of processing 256 lanes per gel per day, which is 5 to 6 times more efficient than methods using large gels, which processes the degree of approximately 32 lanes per day.

[Example 4] Isolation of rice-breed distinguishing bands identified by genomic scanning and designing of specific primers by sequencing

The present inventors developed SCAR markers for easy identification of breeds among commercially available rice and built a system with which anyone can readily identify rice breeds by promptly performing PCR.

Bands that distinguish breeds among 10 major commercially available rices were searched for with genomic scanning (AFLP) using 55 primers. The search took only 2 days. Among several bands found suitable for breed identification, a band specific to the breed "Akitakomachi" was provided for electrophoresis in a wide lane. After electrophoresis was completed, the lane was stained with a fluorescent dye (vistra green), cut out, crushed and extracted in TE buffer, and provided for PCR amplification using primers for AFLP. The PCR products were inserted into appropriate plasmids for introduction into E. Coli. After cultivation and amplification of E. Coli, the plasmids were

obtained and the presence of the insert of interest was confirmed with restriction endonuclease. Then the plasmids were sequenced to design primers which specifically amplifies the band of interest (Fig. 9A). When PCR amplification was performed using DNAs from major 10 breeds as templates and using these primers, a specific band was observed to be amplified only in Akitakomachi (Fig. 9B).

[Example 5] Selection of a clone of a library corresponding to a nucleic acid marker

To identify a clone that correspond to a particular band obtained from a genomic library by this method, SCAR markers obtained as in above Example 3 are usually used to select positive clones by performing colony hybridization on high-density membranes carrying the library. However, this approach takes labor and time for a large number of bands. Moreover, inside sequences in the isolated band are not necessarily unique.

A clone in a library corresponding to a particular band may be selected directly by genomic scanning as follows, without isolating marker bands: First, DNAs of genome clones, such as BAC, are extracted from the whole library clones with a plasmid extractor to prepare plates of the same DNA sequences as of plates of the original clones. The whole genome library is then divided into several sublibraries of 1 genome equivalent or smaller size. Thus, each sublibrary contains average 1 clone corresponding to a particular band. Further, using row, column, and plate numbers as coordinates, 10 ng DNAs each is collected from clones on several microplates constituting a sublibrary, which clones have the same coordinate numbers, to obtain coordinate samples corresponding to different coordinate positions. For example, the coordinate sample for row 3 is obtained by collecting 10 ng DNAs each from all clones on row 3, regardless of plate and column numbers. All coordinate samples are prepared in the same manner. The genome scanning is performed in the same manner on these coordinate samples as well as on the whole genome sample as control. Coordinate sample DNAs may be prepared without extracting DNAs of all clones in the genomic library; instead, clones may be cultivated and increased to about 2 ml, mixed in each of rows, columns, or plates to provide

coordinate samples, from which mixture DNAs are extracted. Once a band of interest on the lane amplified from the control whole genome finds corresponding row, column, and plate bands, the numbers represent a 3-dimentional coordinate for the target clone, which enables to pick up the clone from the sublibrary. In case the sublibrary contains several candidate clones, the candidates may be picked up and provided again for genome scanning along with control for final determination (Fig. 10).

This method is particularly powerful for constructing a contig for entire genome by matching all bands obtained through genome scanning with library clones so that all constituting clones of a genomic library are covered. Thus, a whole genome contig may be readily constructed by one person.

Fig. 10A shows how a particular clone is identified from a genomic sublibrary consisting of 6 half-plates (3 full-plates), where the genomic library (average 120 kB, 6 genome equivalents) of *P. oryzae* Cavara (genome size: 40 MB) is divided into 6 genomic sublibraries such that each sublibrary is approximately 1 genome equivalent. In each genomic sublibrary, the coordinate sample representing row 1 was prepared by collecting 10 μg DNAs from all clones on row 1 on 6 half plates (6 columns x 6 half plates = 36 clones). Therefore, each genomic sublibrary contained 20 coordinate samples in total representing 8 rows, 6 columns, and 6 plates, which would identify 8 x 6 x 6 = 288 clones in the entire genomic sublibrary.

As shown in Fig. 10B, electrophoresis was performed by genomic scanning with 22 lanes at a time, where these coordinate samples were used in 20 lanes as templates and control whole genome was used in 2 lanes as a template. Thus, clones corresponding to the bands obtained in the control were readily identified from the sub genomic library.

Since 3 sub genomic libraries may be placed on 1 electrophoresis gel (66 lanes), only 2 gels are needed to cover the entire genome, i.e., 6 sublibraries. Thus, with 1 AFLP cycle with 1 primer pair, search for clones corresponding to approximately 25 to 40 bands over the whole genome library, which is 6 genome equivalent, is completed with 2 gels. Since the standard genomic scanning processes 4 gels per cycle, clones corresponding to 50 to 80 bands obtained by 2 primer

pairs may be identified from the whole genomic library in 1 electrophoresis cycle.

Thus, even in conservative estimation, approximately 50 to 80 bands from 2 primer sets are matched with corresponding clones in 1 electrophoresis cycle; approximately 1,200 to 2,000 bands are matched with corresponding genome in 25 electrophoresis cycle. Given that the genome size of *P. oryzae* Cavara is 40 MB, obtaining 1,500 bands would yield an average band density of 40 MB / 1,500 bands ≒ 27 kB/band in the genome, which means that average 4.4 bands would be obtained in an average clone on 120 KB. Taking the average 6-fold clone redundancy into consideration, this density should be sufficient for constructing a genome contig. Thus, where BAC plasmids are already available, a contig for a whole genomic library consisting of 6 genome equivalents is completed in approximately 1 month.

Plasmids of an amount for 18 plates may be prepared in about 2 weeks given that an automated plasmid extractor is fully available.

Industrial Applicability

The method and electrophoresis apparatus of this invention has made it possible to very easily and efficiently detect, identify, and obtain nucleic acid markers for purposes such as follows:

(1) Development of polymorphism markers to mark a single gene controlling important functions and/or characters by utilizing polymorphism of nucleic acids proximal to the gene.

For this purpose, the organism with the gene of interest may not have a genome map consisting of already-known markers.

(2) Identification and isolation of clones proximal to the target gene in positional cloning

To isolate and clone a single gene controlling important functions and/or characters by positional cloning based only on positional information on the chromosome, markers immediately close to the gene are searched for, and markers for picking up clones in the region around the gene from a genome library such as that of a BAC (bacterial artificial chromosome) are provided. For this purpose, the organism with the gene of interest may not have a genome map consisting of already-known markers.

(3) Gene analysis and high-density mapping

A large number of nucleic acid markers for gene mapping of an organism are provided efficiently. Linkage analysis of F2 or RI lines of the organism and a high-density mapping are also performed at high efficiency. Thus, analysis of quantitative trait loci (QTL) based on contribution of 2 or more genes is also readily performed.

(4) Identification of breeds

Marker bands for breed identification of commercially available foods and seeds, including polished rice, are detected efficiently. Further, obtained bands are isolated and cloned for sequencing in order to design primer for SCAR (Sequence Characterized Amplified Region) analysis. Use of such SCAR markers may allow prompt and rapid breed-identification. Proximal markers to a particular gene may also be rendered as SCAR markers so that they may be used as more easy-to-use markers for breeding and so on or be used to pick up BAC clones near the gene.

(5) Construction a contig covering the whole genome of an organism

Clones composing a genomic library of an organism may be readily linked together to construct a contig covering the whole genome as follows: A genomic library of several genome equivalents is divided into several sublibraries of approximately 1 genome equivalent each. For each of microplate group composing each sublibrary, the rows, columns and plates are designated as x, y, and z axes, respectively. All clone DNAs orthogonal to an axis are grouped together to provide a coordinate sample used as template. These coordinate samples and whole genome DNA as control are placed on an electrophoresis gel as templates and processed by the genomic scanning. Thus, a clone corresponding to bands on the whole genome lanes is identified (Fig. 10). Given that bands are obtained at the density of equivalent to 1 band / 20-50 kB, a whole genome contig, which has a clone redundancy of several folds in average, is completed.